SCHOOL OF INFORMATICS AND COMPUTING

DEPARTMENT OF BIOHEALTH INFORMATICS
Indiana University–Purdue University
Indianapolis

INFO B429
Machine Learning in Bioinformatics

Department of BioHealth Informatics
Indiana University School of Informatics and Computing, Indianapolis
Spring 2017

Course Details	3 credit hours | Room: IT 357 | Wednesdays, 3–5:40 pm
Instructor	Xiaowen Liu, Ph.D., Assistant Professor
Office	719 Indiana Avenue, WK 304
Email / Phone	xwliu@iupui.edu (317) 278-7613
Office Hours	1:30 – 2:30 pm Wednesdays or by appointment
Structure:	3:00 – 4:15 Lecture
	4:15 – 4:30 Break
	4:30 – 5:40 Lecture

Prerequisites: INFO B419 Introduction to Bioinformatics and either INFO I421 Applications of Data Mining or CSCI 48100 Data Mining.

COURSE DESCRIPTION

This course covers machine learning theories and methods and their application to biological sequence analysis, gene expression data analysis, genomics and proteomics data analysis, and other problems in bioinformatics.

EXTENDED COURSE DESCRIPTION

Machine learning is a key technology in bioinformatics, especially in the analysis of massive biological datasets. This course covers various applications of machine learning methods to bioinformatics, such as gene expression analysis, protein identification, and motif finding. It provides students with a hand-on learning environment by closely integrating machine learning methods and bioinformatics applications. The topics of the course include genomics and proteomics data analysis using Bayesian decision theory and decision trees; gene expression data analysis using linear classification, logistic regression, clustering, and biclustering; and biological sequence analysis using expectation-maximization and hidden Markov models. The course also covers techniques for obtaining public-domain biological datasets and analyzing and visualizing biological data using R packages for machine learning.

REFERENCE BOOKS

Most teaching contents are covered by lecture slides. Please refer to the weekly schedule below for details about the contents that will be covered in the lectures. Students are encouraged to get a copy of the reference books.

Title:	*Bioinformatics: The Machine Learning Approach*
Authors:	Pierre Baldi and Soren Brunak
Edition:	2nd edition (2001)
Publisher:	Bradford Book
ISBN:	026202506X

Title:	*Learn from Data*
Authors:	Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin
Edition:	1st edition (March 2012)
Publisher:	AMLBook
ISBN:	1600490069

GENERAL GUIDELINE TO THE SYLLABUS

Students are responsible for familiarizing themselves with the syllabus. The instructor is responsible for being responsive to the diverse needs of the enrolled students and for making necessary modifications to this syllabus, which is to be treated as a living document.

Software used: R

Teaching and Learning Methods

Project-based learning (PBL), Team-based learning, Lecture by instructor with slides.

LEARNING OUTCOMES:

Upon completion of this course, students will	RBT	PUL	Assessment
1. Access public-domain biological datasets.	3	1C	A1, Pproject
2. Analyze genomics and proteomics data using decision theories, decision trees, and random forests.	4	1B, 3	A1, A3, P, M, F
3. Analyze gene expression data using linear classification, logistic regression, SVM, clustering, and biclustering.	4	1B, 3	A2, A3, P, M, F
4. Analyze biological sequence data using expectation-maximization methods and hidden Markov models.	4	1B, 3	A3, A4, P, F
5. Analyze and visualize biological data sets using R packages for machine learning.	4	1A, 1B	A1, A2, A3, A4, P
6. Design computational experiments for training and evaluating machine learning methods for solving bioinformatics problems.	6	1B, 2	P

Principles of Undergraduate Learning (PUL):

Learning outcomes are assessed in the following areas:

1A. Core communication: written, oral and visual skills

1B. Core communication: quantitative skills   Major emphasis

1C. Core communication: information resources skills

2. Critical thinking   Some emphasis

3. Integration and application of knowledge   Moderate emphasis

4. Intellectual depth, breadth, and adaptiveness

5. Understanding society and culture

6. Values and ethics

Lectures

Lecture 1 Introduction to machine learning

Lecture 2 Linear models for gene expression data analysis

Lecture 3 Decision theory for protein identification

Lecture 4 VC dimensions and parameter estimation

Lecture 5 Logistic regression and neural networks for gene expression data analysis

Lecture 6 Clustering and biclustering for gene expression data analysis

Lecture 7 Expectation maximization methods for sequence analysis

Lecture 8 Decision trees for omics data analysis

Lecture 9 Support vector machines for gene expression data analysis

Lecture 10 Hidden Markov models for sequence analysis

Lecture 11 Dimensionality reduction

Lecture 12 Ensemble learning

ASSIGNMENTS, PROJECT and EXAMS:

Four home assignments will be given to assess learning and apprehension. Homework assignments will be available on Canvas and students will have two weeks to complete each assignment. Scores and model answers will be available on Canvas.

Students will complete a project about compare and evaluate computational methods of biomedical data analysis. Students will work in a group of 2-4 students. Each group will find a biomedical data analysis problem, understand various computational methods, design and perform experiments for method comparison and evaluation. and report analysis reports. Each group will submit a methodology report and a final report, and give a 20-minute oral presentation. Project reports and presentations will be graded based on research objective setting, research redesign, technical complexity, implementation, and experimental results.

A mid-term exam and a final exam will be given to assess student learning outcomes. The mid-

term exam will be in the 8th week and the final exam will be in the 16th week. Both will be 2-hours closed book exams. An unexcused absence from an exam will receive zero. Illness or a death in the immediate family is usually the only acceptable excuse for absence. A makeup exam with 10% penalty will be given to students that miss an exam with an acceptable excuse.

COURSE GRADE BREAKDOWN
- Homework   20%
- Attendance   5%
- Exams   50%
  - Mid-term exam   20%
  - Final exam   30%
- Project   25%
  - Methodology report   8%
  - Final report   17%

One point shall be subtracted for each absence of the class from the final score. At most 5 points will be deducted for absences.

Grading Scale:

A+	97–100%	Professional level work, showing highest level of achievement
A	93–96.99%	Extraordinarily high achievement, quality of work; shows command of the subject matter
A–	90–92.99%	Excellent and thorough knowledge of the subject matter
B+	87–89.99%	Above average understanding of material and quality of work
B	83–86.99%	Mastery and fulfillment of all course requirements; good, acceptable work
B–	80–82.99%	Satisfactory quality of work
C+	77–79.99%	Modestly acceptable performance and quality of work
C	73–76.99%	Minimally acceptable performance and quality of work
C–	70–72.99%	Unacceptable work (Core course must be repeated for credit)
D+	67–69.99%	Unacceptable work (Course must be repeated for credit)
D	63–66.99%	Unacceptable work
D–	60–62.99%	Unacceptable work
F	Below 60	Unacceptable work

No credit is granted for a grade below C.

ATTENDANCE

1. Basic Policy
a. All attendance and assignment deadline policies are in place to protect student educational rights, maintain grading equity, and promote team morale.
b. Attendance shall be taken in every class. If you do not sign the attendance sheet while in

class, you shall be marked absent. Signing the attendance sheet for another student is prohibited.
   c. Illness or a death in the immediate family is usually the only acceptable excuse for absence from class. Absences must be explained to the satisfaction of the instructor.
   d. A student's attendance score (5 points) will be subtracted by 1 point for each absence. At most 5 points will be subtracted for absences.

2. Administrative Withdrawal [University Policy]
   a. A basic requirement of this course is that you will participate in all class meetings and conscientiously complete all required course activities and/or assignments. Keep in touch with the instructor if you are unable to attend, participate, or complete an assignment on time.
   b. If you miss more than half of the required activities within the first 25% of the course without contacting the instructor, you may be administratively withdrawn from this course by the instructor. For example: *This course meets once per week; thus if you miss more than two classes in the first four weeks*, you may be withdrawn by the instructor. Administrative withdrawal may have academic, financial, and financial aid implications. Administrative withdrawal will take place after the full refund period, and if you are administratively withdrawn from the course you will not be eligible for a tuition refund.
   c. If you have questions about the administrative withdrawal policy at any point during the semester, please contact the instructor. See campus policy in detail here: http://registrar.iupui.edu/withdrawal-policy.html

ASSIGNMENT DEADLINES

1. **Late Assignments**

   You are responsible for completing each deliverable (e.g., assignment, project reports) by its deadline and submitting it by the specified method. Deadlines are outlined in the syllabus or in supplementary documents accessible through Canvas. Should you miss a class, you are still responsible for completing the deliverable and for finding out what was covered in class, including any new or modified deliverable. You are allowed one excused or unexcused late submission for homework assignments in 24 hours after the deadline. The second late submission for homework assignments and all late submissions for project reports will receive a zero.

2. **Team Responsibility**

   If a late assignment is due to the action of one team member, the entire team will reap the negative results. For this reason, it is imperative that team members establish a self-monitoring system that includes regular communication via email, text or phone. If a team has a team member who is not acting responsibly, the team may petition the instructor for a solution.

CODE OF CONDUCT

All students should aspire to the highest standards of academic integrity. Using another student's work on an assignment, cheating on a test, not quoting or citing references correctly, or any other form of dishonesty or plagiarism shall result in a grade of zero on the item and possibly an F in the course. Incidences of academic misconduct shall be referred to the Department Chair and repeated violations shall result in dismissal from the program.

All students are responsible for reading, understanding, and applying the *Code of Student Rights, Responsibilities and Conduct* and in particular the section on academic misconduct. Refer to *The Code > Responsibilities > Academic Misconduct* at http://www.indiana.edu/~code/. All students must also successfully complete the Indiana University Department of Education "How to Recognize Plagiarism" Tutorial and Test. https://www.indiana.edu/~istd You must document the difference between your writing and that of others. Use quotation marks in addition to a citation, page number, and reference whenever writing someone else's words (e.g., following the *Publication Manual of the American Psychological Association*). To detect plagiarism instructors apply a range of methods, including Turnitin.com. http://www.ulib.iupui.edu/libinfo/turnitin

Academic misconduct:

1. **Cheating:** Cheating is an attempt to use or provide unauthorized assistance, materials, information, or study aids in any form and in any academic exercise or environment.
   a. A student must not use external assistance on any "in-class" or "take-home" examination, unless the instructor specifically has authorized external assistance. This prohibition includes, but is not limited to, the use of tutors, books, notes, calculators, computers, and wireless communication devices.
   b. A student must not use another person as a substitute in the taking of an examination or quiz, nor allow other persons to conduct research or to prepare work, without advanced authorization from the instructor to whom the work is being submitted.
   c. A student must not use materials from a commercial term paper company, files of papers prepared by other persons, or submit documents found on the Internet.
   d. A student must not collaborate with other persons on a project and submit a copy of a written report that is represented explicitly or implicitly as the student's individual work.
   e. A student must not use any unauthorized assistance in a laboratory, at a computer terminal, or on fieldwork.
   f. A student must not steal examinations or other course materials, including but not limited to, physical copies and photographic or electronic images.
   g. A student must not submit substantial portions of the same academic work for credit or honors more than once without permission of the instructor or program to whom the work is being submitted.
   h. A student must not, without authorization, alter a grade or score in any way, nor alter answers on a returned exam or assignment for credit.
2. **Fabrication:** A student must not falsify or invent any information or data in an academic exercise including, but not limited to, records or reports, laboratory results, and citation to the sources of information.
3. **Plagiarism:** Plagiarism is defined as presenting someone else's work, including the work of other students, as one's own. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged, unless the information is common knowledge. What is considered "common knowledge" may differ from course to course.
   a. A student must not adopt or reproduce ideas, opinions, theories, formulas, graphics, or pictures of another person without acknowledgment.
   b. A student must give credit to the originality of others and acknowledge indebtedness whenever:
      1. directly quoting another person's actual words, whether oral or written;

2. using another person's ideas, opinions, or theories;
3. paraphrasing the words, ideas, opinions, or theories of others, whether oral or written;
4. borrowing facts, statistics, or illustrative material; or
5. offering materials assembled or collected by others in the form of projects or collections without acknowledgment

4. **Interference:** A student must not steal, change, destroy, or impede another student's work, nor should the student unjustly attempt, through a bribe, a promise of favors or threats, to affect any student's grade or the evaluation of academic performance. Impeding another student's work includes, but is not limited to, the theft, defacement, or mutilation of resources so as to deprive others of the information they contain.

5. **Violation of course rules:** A student must not violate course rules established by a department, the course syllabus, verbal or written instructions, or the course materials that are rationally related to the content of the course or to the enhancement of the learning process in the course.

6. **Facilitating academic dishonesty:** A student must not intentionally or knowingly help or attempt to help another student to commit an act of academic misconduct, nor allow another student to use his or her work or resources to commit an act of misconduct.

OTHER POLICIES

1. **Administrative withdrawal:** Students must participate in all class discussions and conscientiously complete all required course activities and/or assignments. If a student is unable to attend, participate in, or complete an assignment on time, the student must inform the instructor. If a student misses more than half of the required activities within the first 25% of the course without contacting the instructor, the student may be administratively withdrawn from this course. Administrative withdrawal may have academic, financial, and financial aid implications. Administrative withdrawal occurs after the full refund period, and a student who has been administratively withdrawn is ineligible for a tuition refund.

2. **Civility:** To maintain an effective and inclusive learning environment, it is important to be an attentive and respectful participant in lectures, discussions, group work, and other classroom exercises. Thus, unnecessary disruptions should be avoided, such as ringing cell phones, engagement in private conversations, and other unrelated activities. Cell phones, media players, or any noisy devices should be turned off during a class. Texting, web surfing, and posting to social media are generally not permitted. Laptop use may be permitted if it is used for taking notes or conducting class activities. Students should check with the instructor about permissible devices in class. IUPUI nurtures and promotes "a campus climate that seeks, values, and cultivates diversity in all of its forms and that provides conditions necessary for all campus community members to feel welcomed, supported, included, and valued" (IUPUI Strategic Initiative 9). IUPUI prohibits "discrimination against anyone for reasons of race, color, religion, national origin, sex, sexual orientation, marital status, age, disability, or veteran status" (Office of Equal Opportunity). Profanity or derogatory comments about the instructor, fellow students, invited speakers or other classroom visitors, or any members of the campus community shall not be tolerated. A violation of this rule shall result in a warning and, if the offense continues, possible disciplinary action.

3. **Communication:** For classroom-based courses, the instructor or teaching assistant should

respond to emails by the end of the next class or, for online courses, within two Indiana University working days, which excludes weekends and holidays. The instructor should provide weekly office hours or accept appointments for face-to-face, telephone, or teleconferenced meetings, and announce periods of extended absence in advance.

4. **Counseling and Psychological Services (CAPS):** Students seeking counseling or other psychological services should contact the CAPS office at 274-2548 or capsindy@iupui.edu. For more information visit http://life.iupui.edu/caps/.

5. **Course evaluations:** Course evaluations provide vital information for improving the quality of courses and programs. Students are urged to complete one course and instructor evaluation for each section in which they are enrolled at the School of Informatics and Computing with the following exceptions: (a) The student has withdrawn from the course; (b) fewer than five students are enrolled in the section (in which case maintaining anonymity is difficult); and (c) the section is a laboratory that must be taken with a course having a different section number. Course evaluations are completed at https://soic.iupui.edu/app/course-eval/. Course evaluations are typically open from the eleventh week. Course evaluations are anonymous, which means that no one can view the name of the student completing the evaluation. In addition, no one can view the evaluation itself until after the instructor has submitted the final grades. In small sections, demographic information should be left blank, if it could be used to identify the student.

6. **Disabilities policy:** All qualified students enrolled in this course are entitled to reasonable accommodations for a disability. Notify the instructor during the first week of class of accommodations needed. Students requiring accommodations register with Adaptive Educational Services (AES) and complete the appropriate AES-issued before receiving accommodations. The AES office is located at UC 100, Taylor Hall (Email: aes@iupui.edu, Tel. 317 274-3241). For more information visit http://aes.iupui.edu.

7. **Email:** Indiana University uses your IU email account as an official means of communication, and students should check it daily. Although you may have your IU email forwarded to an outside email account, please email faculty and staff from your IU email account.

8. **Emergency preparedness:** Know what to do in an emergency so that you can protect yourself and others. For more information, visit the emergency management website at http://protect.iu.edu/emergency.

9. **IUPUI course policies:** Several campus policies governing IUPUI courses may be found at the following link: http://registrar.iupui.edu/course_policies.html

10. **No class attendance without enrollment.** Only those who are officially enrolled in this course may attend class unless enrolled as an auditor or making up an Incomplete by prior arrangement with the instructor. This policy does not apply to those assisting a student with a documented disability, serving in an instructional role, or administrative personnel. http://registrar.iupui.edu/official-enrollment-class-attendance.html Children may *not* attend class with their parents, guardians, or childcare providers.

11. **Religious holidays:** Students seeking accommodation for religious observances must submit a request form to the course instructor by the end of the second week of the semester. For information visit http://registrar.iupui.edu/religious.html.

12. **Right to revise:** The instructor reserves the right to make changes to this syllabus as necessary

and, in such an event, will notify students of the changes immediately.

13. **Sexual misconduct:** IU does not tolerate sexual harassment or violence. For more information and resources, visit http://stopsexualviolence.iu.edu/.

14. **Student advocate:** The Student Advocate assists students with personal, financial, and academic issues. The Student Advocate is in the Campus Center, Suite 350, and may also be contacted at 317 274-4431 or studvoc@iupui.edu. For more information visit http://studentaffairs.iupui.edu/advocate.

MISSION STATEMENT

The Mission of IUPUI is to provide for its constituents excellence in

- Teaching and Learning;
- Research, Scholarship, and Creative Activity; and
- Civic Engagement.

With each of these core activities characterized by

- Collaboration within and across disciplines and with the community;
- A commitment to ensuring diversity; and
- Pursuit of best practices.

IUPUI's mission is derived from and aligned with the principal components—Communities of Learning, Responsibilities of Excellence, Accountability and Best Practices—of Indiana University's Strategic Directions Charter.

STATEMENT OF VALUES

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana's capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community, both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to continuous improvement of its programs and services.